UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 10/685,586 | 10/16/2003 | Daben Liu | BBNT-P01-086 | 5146 |

28120 7590 04/02/2007

FISH & NEAVE IP GROUP
ROPES & GRAY LLP
ONE INTERNATIONAL PLACE
BOSTON, MA 02110-2624

| EXAMINER | |
|---|---|
| SIEDLER, DOROTHY S | |
| **ART UNIT** | **PAPER NUMBER** |
| 2626 | |

| SHORTENED STATUTORY PERIOD OF RESPONSE | MAIL DATE | DELIVERY MODE |
|---|---|---|
| 3 MONTHS | 04/02/2007 | PAPER |

**Please find below and/or attached an Office communication concerning this application or proceeding.**

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

| Office Action Summary | Application No. | Applicant(s) | |
|---|---|---|---|
| | 10/685,586 | LIU ET AL. | |
| | Examiner | Art Unit | |
| | Dorothy Sarah Siedler | 2626 | |

*-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --*

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

1)☒ Responsive to communication(s) filed on *16 October 2003*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

4)☒ Claim(s) *1-31* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-31* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

**Application Papers**

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *16 October 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *1-2&04,1-8-07,3-12-07*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

## DETAILED ACTION

This is the initial office action is response to the application filled October 16, 2003.

Claims 1-31 are pending and are considered below.

### *Claim Rejections - 35 USC § 101*

35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 1-22 and 30-31 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

Claims 1,11,18 and 30 fall within a judicial exception as they merely manipulate an abstract idea (mathematical algorithm) without a claimed limitation to a practical application. The claimed method is merely a series of steps to be performed on a computer, which manipulates a mathematical algorithm without any claimed limitation to a practical application.

Claims 2-10, 12-17,19-22 and 31 fail to resolve the deficiencies of claims 1,11,18 and 30, and therefore are rejected under similar grounds, i.e. lacking a claimed limitation to a practical application.

### *Claim Rejections - 35 USC § 102*

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

Claims 1,4-10,11,14-17,18, 21,22 and 30 are rejected under 35 U.S.C. 102(b) as being anticipated by ***Liu*** ("Fast Speaker Change Detection for Broadcast News Transcription and Indexing" 1999).

As per claims 1,11,18 and 30, ***Liu*** discloses a method and device for detecting speaker changes in an input audio stream comprising: a processor, and a memory containing instructions (*the system discloses a fast speaker change detection algorithm. Since these algorithms are performed are on a computer, it is inherent that the system uses a processor and memory storing instructions to be executed by the processor*) that when executed by the processor cause the processor to: segment the input audio stream into predetermined length intervals (Section 4 Speaker Change Detection, first paragraph, *the speech is segmented into uniform-length segments*); decode the intervals to produce a set of phones, or phone classes, corresponding to each of the intervals (section 4 Speaker Change Detection, second paragraph, *the speaker change algorithm uses the phone/non-speech sequence produced by the phone-class decode*), a number of possible phone classes being approximately seven (Abstract, *4 broad phoneme classes and 4 non-speech classes*); generate a similarity measurement based on a first portion of the audio stream within one of the intervals and prior to a boundary between adjacent phones and a second portion of the audio stream within the one of the intervals after the boundary, and detecting speaker changes based on the similarity measurement (Section 4 Speaker Change Detection, subsection Phone-based speaker

change detection, *speaker changes are determined on each boundary, segments compared using the distance measure (similarity measure) criterion*).

As per claim 4, ***Liu*** discloses the method of claim 1, wherein generating a similarity measurement includes: calculating cepstral vectors for the audio stream prior to the boundary and the audio stream after the boundary, and comparing the cepstral vectors (Section 4 Speaker Change Detection, subsection Distance Measure Criterion, *cepstral vectors are used in the distance measure (similarity measure)*).

As per claim 5, ***Liu*** discloses the method of claim 4, wherein the cepstral vectors are compared using a generalized likelihood ratio test (Section 4 Speaker Change Detection, subsection Distance Measure Criterion).

As per claim 6, ***Liu*** discloses the method of claim 5, wherein a speaker change is detected when the generalized likelihood ratio test produces a value less than a preset threshold (Section 4 Speaker Change Detection, subsection the critical region).

As per claims 7 and 14, ***Liu*** discloses the method and device of claims 1 and 11, wherein the decoded set of phones is selected from a simplified corpus of phone classes (Section 4 Speaker Change Detection, second paragraph and Section 3 Phone-Class Decode, *the speaker change detection algorithm uses the phone/non-speech*

*sequence produced by the phone-class decode; the phone-class decode section using 4 broad phoneme classes*).

As per claims 8,15 and 21, ***Liu*** discloses the method and device of claims 7,14 and 18, wherein the simplified corpus of phone classes includes a phone class for vowels and nasals, a phone class for fricatives, and a phone class for obstruents (Section 3 Phone-Class Decode, Figure 1).

As per claims 9, 16, and 22, ***Liu*** discloses the method and device of claims 8,15 and 21 wherein the simplified corpus of phone classes further includes a phone class for music, laughter, breath and lip-smack, and silence (Section 4 Speaker Change Detection, second paragraph and Section 3 Phone-Class Decode, *the speaker change detection algorithm uses the phone/non-speech sequence produced by the phone-class decode; the phone-class decode section using 4 non-speech classes,* and Section 2 second paragraph).

As per claims 10 and 17, ***Liu*** discloses the method and device of claims 7, wherein the simplified corpus of phone classes includes approximately seven phone classes (Abstract, *4 broad phoneme classes and 4 non-speech classes*).

***Claim Rejections - 35 USC § 103***

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 2,12, and 19 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Liu*** in view of ***Beigi*** ("A Distance Measure Between Collection of Distributions and it's Application to Speaker Recognition" IEEE 1998).

***Liu*** discloses the method of claims 1,11 and 18, however ***Liu*** does not disclose wherein the predetermined length intervals are approximately thirty seconds in length. ***Beigi*** discloses the use of intervals that are approximately thirty seconds long in a speech recognition system (page 756, Section 4 Results, *30 seconds of speech is used for training*). Therefore, the examiner argues that it is old and well known to segment audio into predetermined intervals approximately 30 seconds long, as indicated by ***Beigi***.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to use predetermined length intervals approximately thirty seconds in length in ***Liu***, since an interval of that length would provide robust data for training a speaker segmentation model.

Claims 3,13,20 and 31 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Liu*** in view of ***Braida*** (6,317,716).

***Liu*** discloses the method of claims 1,11,18 and 30, but does not explicitly disclose wherein segmenting the input audio stream includes: creating the predetermined length intervals such that portions of the intervals overlap one another. ***Braida*** discloses a system that uses overlapping frames (column 7 lines 7-15). Therefore, the examiner argues that it is old and well known to create intervals that overlap one another, as indicated by ***Braida***.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to create overlapping intervals in ***Liu***, since overlapping segments would minimize segmentation errors by ensuring that speaker boundaries are determined at word boundaries instead of in the middle of words.

Claims 23-25, and 27-29 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Liu*** in view of ***Colbath*** ("Spoken Documents: Creating Searchable Archives from Continuous Audio" 2000) further in view of ***Braida***.

As per claim 23, ***Liu*** discloses a system comprising: a segmentation component configured to divide the audio data into segments (Section 4 Speaker Change Detection, first paragraph, *the speech is segmented into uniform-length segments*), a

speaker change detection component configured to detect locations of speaker changes in the audio data based on a similarity value calculated at locations in the segments that correspond to phone class boundaries (Section 4 Speaker Change Detection, subsection Phone-based speaker change detection, *speaker changes are determined on each boundary, segments compared using the distance measure criterion*). However, ***Liu*** does not disclose an indexer configured to receive input audio data and generate a rich transcription from the audio data, the rich transcription including metadata that defines speaker changes in the audio data, the indexer including: a memory system for storing the rich transcription, and a server configured to receive requests for documents and to respond to the requests by transmitting ones of the rich transcriptions that match the requests. ***Liu*** also does not disclose segmenting audio data into overlapping segments. ***Colbath*** discloses an indexer configured to receive input audio data and generate a rich transcription from the audio data, the rich transcription including metadata that defines speaker changes in the audio data (page 2, Component Technologies, first paragraph and page 4, System Architecture, first paragraph, *the system creates a transcript for use in information retrieval systems, the transcript created using speech recognition and speech segmentation techniques, and therefore includes speaker boundaries (metatdata)*), a memory system for storing the rich transcription, and a server configured to receive requests for documents and to respond to the requests by transmitting ones of the rich transcriptions that match the requests (page 4-5, System Architecture, server and browser). In addition, ***Braida*** discloses a system that uses overlapping frames (column 7 lines 7-15). Therefore,

examiner argues that it is old and well known to create intervals that overlap one another, as indicated by ***Braida***.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have an indexer configured to receive input audio data and generate a rich transcription from the audio data, the rich transcription including metadata that defines speaker changes in the audio data, a memory system for storing the rich transcription, and a server configured to receive requests for documents and respond to the requests by transmitting one or more of the rich transcriptions that match the requests in ***Liu***, since it would create a system that integrates acoustic and linguistic technologies to construct a structural summary of continuous audio that is searchable by content, as indicated in ***Colbath*** (page 2, fourth paragraph).

In addition, it would have been obvious to one of ordinary skill in the art at the time of the invention to create overlapping intervals in ***Liu***, since overlapping segments would minimize segmentation errors by ensuring that speaker boundaries are determined at word boundaries instead of in the middle of words.

As per claim 24, ***Liu*** in view of ***Colbath*** further in view of ***Braida*** discloses the system of claim 23, and ***Colbath*** further discloses wherein the indexer further includes at least one of: a speaker clustering component, a speaker identification component, a name spotting component, and a topic classification component (page 2, Component Technologies).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have an indexer include at least one of: a speaker clustering component, a speaker identification component, a name spotting component, and a topic classification component in ***Liu***, since it would create a system that integrates acoustic and linguistic technologies to construct a structural summary of continuous audio that is searchable by content, as indicated in ***Colbath*** (page 2, fourth paragraph).

As per claim 25, ***Liu*** in view of ***Colbath*** further in view of ***Braida*** discloses the system of claim 23, and ***Liu*** further discloses wherein the overlapping segments are segments of a predetermined length (Section 4 Speaker Change Detection, first paragraph, *the speech is segmented into uniform-length segments*).

As per claim 27, ***Liu*** in view of ***Colbath*** further in view of ***Braida*** discloses the system of claim 23, and ***Liu*** further wherein the phone classes include a phone class for vowels and nasals, a phone class for fricatives, and a phone class for obstruents (Section 3 Phone-Class Decode, Figure 1).

As per claim 28, ***Liu*** in view of ***Colbath*** further in view of ***Braida*** discloses the system of claim 27, and ***Liu*** further discloses wherein the phone classes additionally include a phone class for music, laughter, breath and lip-smack, and silence (Section 4 Speaker Change Detection, second paragraph and Section 3 Phone-Class Decode, *the speaker*

*change detection algorithm uses the phone/non-speech sequence produced by the phone-class decode; the phone-class decode section using 4 non-speech classes*).

As per claim 29, ***Liu*** in view of ***Colbath*** further in view of ***Braida*** discloses the system of claim 23, and ***Liu*** further discloses wherein the phone classes include approximately seven phone classes (Abstract, *4 broad phoneme classes and 4 non-speech classes*).

Claim 26 is rejected under 35 U.S.C. 103(a) as being unpatentable over ***Liu*** in view of ***Colbath*** ("Spoken Documents: Creating Searchable Archives from Continuous Audio" 2000) further in view of ***Braida*** and further in view of ***Beigi***.

***Liu*** in view of ***Colbath*** further in view of ***Braida*** discloses the system of claim 25, however neither ***Liu***, ***Colbath*** nor ***Braida*** discloses wherein the predetermined length is approximately thirty seconds. ***Beigi*** discloses the use of intervals that are approximately thirty seconds long in a speech recognition system page 756, Section 4 Results, *30 seconds of speech is used for training*). Therefore, examiner argues that it is old and well known to segment audio into predetermined intervals approximately 30 seconds long, as indicated by ***Beigi***.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to use predetermined length intervals approximately thirty seconds in

length in ***Liu***, since an interval of that length would provide robust data for training a speaker segmentation model.

### *Conclusion*

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dorothy Sarah Siedler whose telephone number is 571-270-1067. The examiner can normally be reached on Mon-Thur 9:30am-5:30pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on 571-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

DSS

TĀLIVALDIS IVARS ŠMITS
PRIMARY EXAMINER